

ABSTRAK

Penyakit jantung merupakan salah satu penyakit tidak menular yang dapat menyebabkan kematian. Penyakit ini perlu ditangani serius dengan identifikasi gejala secara dini. Diagnosa penyakit jantung sering kali dibuat dengan berfokus pada naluri dan pengalaman dokter, bukan berdasarkan informasi yang yang diperoleh dari *database*. Hal ini menimbulkan penilaian bias yang tidak diinginkan, kesalahan dan biaya pengobatan yang tidak perlu, sehingga mempengaruhi kualitas layanan yang diberikan kepada pasien. Di pusat medis (rumah sakit atau pusat klinis lainnya) teknik data mining membantu merawat pasien untuk mengidentifikasi apakah pasien tersebut memiliki penyakit atau tidak, serta dapat mendiagnosa gejala penyakit jantung secara dini pada pasien dalam waktu singkat. Pada penelitian ini akan menggunakan metode *Logistics Regression Ensemble (LORENS)* untuk mengklasifikasi gejala penyakit jantung dan juga menghitung tingkat akurasi dari penerapan metode tersebut. Metode ini dipilih untuk mengetahui apakah metode *LORENS* mampu diterapkan pada dataset penyakit jantung dimana dataset tersebut mempunyai dimensi data rendah. Hasil penelitian ini menunjukkan nilai akurasi sebesar 83,63% jika menerapkan *LORENS* dengan *One Hot Encoding*, sedangkan jika menerapkan *LORENS* tanpa *One Hot Encoding* nilai akurasi sebesar 81,65 %. Nilai akurasi yang pertama lebih besar karena pada dataset terdapat data kategorikal sehingga dilakukan proses *One Hot Encoding* untuk mentransformasi data kategorikal ke numerik. Berdasarkan hasil tersebut dapat disimpulkan metode *LORENS* dapat digunakan untuk klasifikasi penyakit jantung.

Kata Kunci : *Ensemble Learning, Logistic Regression, Data Mining, Heart disease.*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
KATA PENGATAR	iv
ABSTRAK	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.2.1 Permasalahan	3
1.2.2 Metode Usulan.....	3
1.2.3 Pertanyaan Penelitian.....	3
1.3 Tujuan dan Manfaat	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	4
1.4 Batasan-batasan.....	4
1.5 Sistematika Penulisan.....	5
BAB II KAJIAN PUSTAKA	6
2.1 <i>Data Mining</i>	6
2.2 Penyakit Jantung	6
2.3 <i>Logistic Regression</i>	6
2.4 <i>Logistic Regression Ensemble</i> (LORENS)	7
2.5 <i>Ensemble Vote</i>	9
2.6 Normalisasi <i>Min-Max</i>	10
2.7 <i>One Hot Encoding</i>	10
2.8 Evaluasi	11
2.8.1 Metode Evaluasi.....	11
2.8.2 Matriks Evaluasi	11
2.8 Penelitian Terkait	12
BAB III METODE USULAN	16
3.1 Metode Penelitian.....	16
3.1.1 Studi Pustaka.....	16

3.1.2 Perancangan Sistem	16
3.1.3 Implementasi Sistem.....	18
3.1.4 Pengujian Sistem.....	18
3.1.5 Kesimpulan dan Saran	19
3.2 Flowchart Sistem.....	19
3.3 Perhitungan Manual	24
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Lingkungan Uji Coba	37
4.2 Dataset.....	37
4.3 Implementasi Metode	40
4.3.1 Implementasi Metode <i>One Hot Encoding</i>	40
4.3.2 Implementasi Metode <i>Min-Max</i>	41
4.3.3 Implementasi Metode <i>Logistic Regression</i>	41
4.3.4 Implementasi Metode <i>Logistic Regression Ensemble (LORENS)</i>	42
4.3.5 Implementasi Metode <i>K-Fold Cross Validation</i>	46
4.4 Implementasi <i>User Interface</i>	49
4.4.1 Tampilan <i>Dataset</i> awal dan <i>One Hot Encoding</i>	49
4.4.2 Tampilan Normalisasi <i>Min-Max</i>	50
4.4.3 Tampilan membuat partisi acak pada metode <i>LORENS</i>	50
4.4.4 Tampilan hasil klasifikasi	51
4.4.5 Tampilan detail hasil klasifikasi	52
4.4.6 Tampilan Menghitung <i>mean</i> setiap <i>ensemble</i>	52
4.4.7 Tampilan <i>input</i> data baru	53
4.5 Implementasi Klasifikasi.....	54
4.5.1 Skenario Pengujian Pertama	54
4.5.2 Skenario Pengujian Kedua	58
4.5.3 Evaluasi.....	61
BAB V PENUTUP	64
5.1 Kesimpulan.....	64
5.2 Saran.....	65
DAFTAR PUSTAKA	66
LAMPIRAN.....	68

DAFTAR GAMBAR

Gambar 2.1 <i>Binary Logistic Regression</i>	7
Gambar 2.2 Diagram <i>Ensemble Vote</i>	9
Gambar 2.3 Metode evaluasi <i>K-Fold Cross Validation</i>	11
Gambar 3.1 Bagan Penelitian	16
Gambar 3.2 <i>Flowchart</i> LORENS	19
Gambar 3.3 <i>Flowchart</i> Koefisien	20
Gambar 3.4 <i>Flowchart</i> Klasifikasi	21
Gambar 3.5 <i>Flowchart One Hot Encoding</i>	22
Gambar 3.6 <i>Flowchart Min-Max</i>	23
Gambar 4.1 Tampilan <i>Dataset</i> awal dan <i>One Hot Encoding</i>	49
Gambar 4.2 Tampilan Normalisasi <i>Min-Max</i>	50
Gambar 4. 3 Tampilan membuat Partisi acak pada metode LORENS	50
Gambar 4.4 Tampilan hasil klasifikasi	51
Gambar 4.5 Tampilan detail hasil klasifikasi	52
Gambar 4.6 Tampilan menghitung <i>mean</i> setiap <i>ensemble</i>	52
Gambar 4.7 Tampilan <i>input</i> data baru	53
Gambar 4.8 Tampilan hasil klasifikasi <i>input</i> data baru	53
Gambar 4. 9 Pengujian pertama - hasil akurasi	56
Gambar 4. 10 Pengujian pertama - hasil <i>precision</i>	57
Gambar 4. 11 Pengujian pertama - hasil <i>recall</i>	57
Gambar 4. 12 Pengujian pertama - hasil <i>mean</i>	57
Gambar 4. 13 Pengujian kedua - hasil akurasi	59
Gambar 4. 14 Pengujian kedua - hasil <i>precision</i>	60
Gambar 4. 15 Pengujian kedua - hasil <i>recall</i>	60
Gambar 4. 16 Pengujian kedua - hasil <i>mean</i>	60
Gambar 4. 17 Pengujian pertama - hasil kinerja <i>Logistics Regression</i>	62
Gambar 4.18 Pengujian kedua - hasil kinerja <i>Logistics Regression</i>	62
Gambar 4.19 Pengujian pertama dan kedua - hasil <i>Mean Logistics Regression</i>	62

DAFTAR TABEL

Tabel 2.1 Sebelum di <i>One Hot Encoding</i>	10
Tabel 2.2 Setelah di <i>One Hot Encoding</i>	11
Tabel 2.3 <i>Confusion Matrix</i>	12
Tabel 2.4 Penelitian terkait	14
Tabel 3.1 <i>Dataset</i> awal.....	24
Tabel 3.2 Hasil <i>One Hot Encoding</i>	24
Tabel 3.3 Hasil <i>Min-Max</i>	25
Tabel 3.4 Data latih	26
Tabel 3.5 Data uji	26
Tabel 3.6 Partisi acak	27
Tabel 3.7 Tambah bias pada data latih di <i>Ensemble_1</i> - partisi 1	27
Tabel 3.8 Hasil koefisien pada <i>Ensemble</i> di setiap partisi	29
Tabel 3.9 Tambah bias pada data uji di <i>Ensemble_1</i> - partisi 1	30
Tabel 3.10 Tambah bias pada data uji di <i>Ensemble_1</i> - partisi 2	30
Tabel 3.11 Tambah bias pada data uji di <i>Ensemble_1</i> - partisi 3	30
Tabel 3.12 hasil klasifikasi pada <i>Ensemble_1</i>	33
Tabel 3.13 hasil klasifikasi pada <i>Ensemble_2</i>	33
Tabel 3.14 hasil klasifikasi pada <i>Ensemble_3</i>	33
Tabel 3.15 kelas aktual dan kelas prediksi <i>Ensemble_1</i>	34
Tabel 3.16 kelas aktual dan kelas prediksi <i>Ensemble_2</i>	34
Tabel 3.17 kelas aktual dan kelas prediksi <i>Ensemble_3</i>	34
Tabel 3.18 <i>Confusion Matrix</i> <i>Ensemble_1</i>	35
Tabel 3.19 <i>Confusion Matrix</i> <i>Ensemble_2</i>	35
Tabel 3.20 <i>Confusion Matrix</i> <i>Ensemble_3</i>	35
Tabel 4.1 Spesifikasi Perangkat	37
Tabel 4.2 Keterangan <i>Dataset</i>	37
Tabel 4. 3 keterangan dataset setelah di <i>One Hot Encoding</i>	38
Tabel 4.4 Tabel memanggil <i>Library pandas, numpy, math</i>	40
Tabel 4.5 Kode Program <i>One Hot Encoding</i>	40
Tabel 4.6 kode program <i>Min-Max</i>	41
Tabel 4.7 Kode program menghitung koefisien.....	41
Tabel 4.8 Kode program klasifikasi	42
Tabel 4.9 Kode program membuat partisi acak.....	43
Tabel 4.10 Kode program <i>Majority Voting</i>	43
Tabel 4.11 Kode program <i>Confusion Matrix</i>	44
Tabel 4.12 Kode program <i>Logistic Regression Ensemble (LORENS)</i>	44
Tabel 4.13 Kode program <i>K-fold Cross Validation</i>	46
Tabel 4.14 Pengujian pertama - fitur <i>dataset</i> awal.....	54
Tabel 4.15 Pengujian pertama - fitur setelah <i>On Hot Encoding</i>	54
Tabel 4.16 Pengujian pertama - Hasil <i>One Hot Encoding</i>	54
Tabel 4.17 Pengujian pertama - hasil <i>Min-Max</i>	55
Tabel 4.18 Pengujian pertama - partisi acak	55

Tabel 4.19 Pengujian pertama - hasil <i>threshold</i> setiap <i>fold</i>	56
Tabel 4.20 Pengujian kedua - fitur dataset awal.....	58
Tabel 4. 21 Pengujian kedua - hasil <i>Min-Max</i>	58
Tabel 4.22 Pengujian kedua - partisi acak	59
Tabel 4.23 Pengujian kedua - hasil <i>threshold</i> setiap <i>fold</i>	59